



**The effect of incremental changes in phonotactic probability
and neighborhood density on word learning by preschool
children**

Journal:	<i>Journal of Speech, Language, and Hearing Research</i>
Manuscript ID:	JSLHR-L-12-0245.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Storkel, Holly; University of Kansas, Speech-Language-Hearing: Sciences and Disorders Bontempo, Daniel; University of Kansas, Bureau of Child Research Aschenbrenner, Andrew; Washington University, Psychology Maekawa, Junko; University of Kansas, Speech-Language-Hearing Sciences and Disorders Lee, SuYeon; Hallym University, Speech pathology and audiology
Keywords:	Language, Phonology, Children

Running Head: INCREMENTAL CHANGES

REGULAR PAPER

The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children

Holly L. Storkel
Daniel E. Bontempo
Andrew J. Aschenbrenner*
Junko Maekawa
Su-Yeon Lee**

University of Kansas

Resubmitted (02/20/2013), *Journal of Speech, Language, and Hearing Research*

*Andrew J. Aschenbrenner is now at Washington University – Saint Louis.
**Su-Yeon Lee is now at Woosong University, Daejeon, South Korea.

Send correspondence to:
Holly L. Storkel
Speech-Language-Hearing: Sciences & Disorders
University of Kansas
1000 Sunnyside Ave
Dole Human Development Center, 3001
Lawrence, KS 66045-7555
784-864-0497
hstorkel@ku.edu

Abstract

Purpose. Phonotactic probability or neighborhood density have predominately been defined using gross distinctions (i.e., low vs. high). The current studies examined the influence of finer changes in probability (Experiment 1) and density (Experiment 2) on word learning.

Method. The full range of probability or density was examined by sampling five nonwords from each of four quartiles. Three- and 5-year-old children received training on nonword-nonobject pairs. Learning was measured in a picture-naming task immediately following training and 1-week after training. Results were analyzed using multi-level modeling.

Results. A linear spline model best captured nonlinearities in phonotactic probability. Specifically word learning improved as probability increased in the lowest quartile, worsened as probability increased in the midlow quartile, and then remained stable and poor in the two highest quartiles. An ordinary linear model sufficiently described neighborhood density. Here, word learning improved as density increased across all quartiles.

Conclusion. Given these different patterns, phonotactic probability and neighborhood density appear to influence different word learning processes. Specifically, phonotactic probability may affect recognition that a sound sequence is an acceptable word in the language and is a novel word for the child, whereas neighborhood density may influence creation of a new representation in long-term memory.

Key words: vocabulary, word learning, phonotactic probability, neighborhood density, spline regression

The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children

Learning is influenced by language structure, including *phonotactic probability*, which is the frequency of occurrence of a sound in a given word position and/or the frequency of co-occurrence of adjacent sound combinations, and *neighborhood density*, which refers to the number of words that differ by one phoneme from a given word (Vitevitch & Luce, 1998, 1999). When probability and density are correlated, children learn high probability/density sound sequences more accurately than low probability/density (Storkel, 2001, 2003, 2004a; Storkel & Maekawa, 2005). When probability and density are differentiated, young children and adults still learn high density sound sequences more accurately than low density, but they now learn low probability sound sequences more accurately than high probability (Hoover, Storkel, & Hogan, 2010; Storkel, 2009; Storkel, Armbruster, & Hogan, 2006; Storkel & Lee, 2011).

Even though there is clear evidence that phonotactic probability and neighborhood density influence learning, the majority of evidence to date has only considered gross distinctions in phonotactic probability and neighborhood density. That is, virtually all empirical studies contrast “low” versus “high” probability or density, even though phonotactic probability and neighborhood density are continuous variables. Consequently, it is unclear at present whether smaller incremental differences in probability and/or density influence word learning. Likewise, the pattern of performance across the distribution of probability and density is unknown. On the one hand, children could show a nonlinear pattern, such that small changes at certain points on the probability or density distribution would improve (or worsen) performance while changes at other points would lead to minimal or no change in performance (e.g., stable performance). On the other hand, children could show a linear pattern, such that even small changes in probability and density would improve (or worsen) performance across the full distribution of probability and density. The overarching goal of this research is to examine whether smaller differences in

1
2
3 probability and/or density influence performance on a word learning task and to determine the
4
5 pattern of performance across the full distribution of probability and density.
6

7
8 There is reason to predict that the pattern of word learning performance across the
9
10 distribution of phonotactic probability will differ from that of neighborhood density (Storkel, et
11
12 al., 2006; Storkel & Lee, 2011). Prior probability and density findings have been interpreted
13
14 within a model of word learning that differentiates three processes: triggering, configuration, and
15
16 engagement (cf. Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Leach & Samuel, 2007; Li,
17
18 Farkas, & Mac Whinney, 2004). *Triggering* involves allocation of a new representation (i.e.,
19
20 recruitment of a new node in a connectionist network), which occurs when the mismatch
21
22 between the input and existing representations exceeds a set threshold (e.g., the vigilance
23
24 parameter, Li, et al., 2004). In this way, a novel input is detected and learning (i.e., recruitment
25
26 of a new node) is initiated. *Configuration* entails the actual creation of the new representation in
27
28 long-term memory (e.g., storing information in the newly allocated representation, Leach &
29
30 Samuel, 2007; Li, et al., 2004). Finally, *engagement* is the integration of a newly created
31
32 representation with similar existing representations in long-term memory, which may require a
33
34 delay that includes sleep (e.g., forming connections between similar representations, Dumay &
35
36 Gaskell, 2007; Leach & Samuel, 2007; Li, et al., 2004).
37
38
39
40
41
42

43
44 Prior word learning research with children and adults suggests that phonotactic
45
46 probability may influence triggering, whereas neighborhood density may influence configuration
47
48 and/or engagement (Storkel, et al., 2006; Storkel & Lee, 2011). This finding also is consistent
49
50 with the DevLex model (Li, et al., 2004) where neighborhood relationships are essentially turned
51
52 off during triggering (i.e., node recruitment) to maintain the stability of previously learned
53
54 words, but neighborhood structure is available during other types of processing (e.g.,
55
56 configuration and engagement). Based on the prior experimental findings and the DevLex model,
57
58 the pattern of word learning performance across the phonotactic probability distribution is
59
60

1
2
3 predicted to differ from the pattern of word learning performance across the neighborhood
4
5 density distribution.
6

7 In terms of specific predictions for triggering and phonotactic probability, past findings
8
9 from disambiguation studies are relevant. In disambiguation studies, children are presented with
10
11 at least one known object and at least one novel object along with a novel sound sequence, and
12
13 looking behavior is measured. Presumably, if the child recognizes that the sound sequence is
14
15 novel, more looks will be directed toward the novel object. This disambiguation is a type of
16
17 triggering (i.e., recognizing novelty). At least a few studies have examined the effect of novelty
18
19 on looking behavior in this paradigm with novelty being defined by the number of feature
20
21 differences between the novel sound sequence and the known name of the real object. In fact,
22
23 looks to the novel object increase as the novelty of the sound sequence increases. However,
24
25 when the novel word is minimally novel, children still overwhelmingly choose or look at the real
26
27 object rather than the novel object (e.g., Creel, 2012; White & Morgan, 2008). This suggests that
28
29 triggering may not be occurring for minimally novel stimuli. In terms of a prediction for the
30
31 current study, word learning performance may decrease (linearly) as phonotactic probability
32
33 increases due to inefficient triggering as the word becomes less novel. Then, stable poor
34
35 performance may be observed at the higher end of the probability distribution where the items
36
37 may no longer be detected as novel. Here, the assumption is that triggering does not occur at all,
38
39 leading to uniformly poor word learning. An additional, as yet unexplored, possibility is that this
40
41 shift in the effect of phonotactic probability (i.e., linear decrease in performance followed by
42
43 stable poor performance) could occur rapidly, leading to a discontinuity in the function relating
44
45 phonotactic probability to word learning performance.
46
47
48
49
50
51
52
53

54 Turning to specific predictions for configuration and neighborhood density, working
55
56 memory theory suggests that word learning performance should linearly increase as
57
58 neighborhood density increases. This is based on the assumption that an item in working
59
60

Incremental Changes 6

memory is supported by the activation of items in long-term memory (Roodenrys & Hinton, 2002; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). Thus, the more items activated (i.e., the higher the density), the greater the support to working memory from long-term memory, with no apparent cap on this support (i.e., no expectation of nonlinearity). Further, it is assumed that the integrity of the item in working memory influences the integrity of the newly stored item in long-term memory, namely configuration (Gathercole, 2006). Predictions from an engagement perspective are somewhat less clear because there has been less research in this area but presumably the logic is somewhat similar to that just described for configuration. That is, connections to existing items in long-term memory provide support to the newly created representation. The more connections created (i.e., the higher the density), the greater the support from existing representation, with no apparent cap on this support (i.e., no expectation of nonlinearity).

The current research makes an initial attempt to address these issues in two word learning experiments with 3- and 5-year-old children. Experiment 1 examined learning of novel words varying in phonotactic probability but matched in neighborhood density. Experiment 2 examined learning of novel words varying in neighborhood density but matched in phonotactic probability. For both experiments, the full range of the distribution of phonotactic probability or neighborhood density was sampled by dividing that distribution into four ranges defined by quartiles, lowest ($< 25^{\text{th}}$ percentile), midlow ($25^{\text{th}} - 49^{\text{th}}$ percentile), midhigh ($50^{\text{th}} - 74^{\text{th}}$ percentile), and highest ($\geq 75^{\text{th}}$ percentile), and sampling five items within each quartile. Based on past word learning research, phonotactic probability and neighborhood density were predicted to have a significant effect on word learning accuracy. The main contribution of this research is examination of the relationship between word learning accuracy and the full distribution of probability or density. Several models were fit to the data to examine a variety of potential patterns including discontinuous and nonlinear patterns as well as continuous linear patterns.

Experiment 1: Phonotactic Probability

Learning of novel words varying in probability (i.e., lowest, midlow, midhigh, highest) but matched in density and nonobject characteristics (i.e., objectlikeness ratings, number of semantic neighbors) was examined.

Method

Participants. Twenty-three 3-year-old ($M = 3$ years; 8 months; $SD = 0; 3$; range = 3; 1 – 3; 11) and 24 5-year-old ($M = 5$ years; 4 months; $SD = 0; 3$; range = 5; 0 – 6; 0) children participated. All children were monolingual native speakers of English with no history of speech, language, motor, cognitive, or health impairment by parent report. Standardized clinical testing (Dunn & Dunn, 2007; Goldman & Fristoe, 2000; Williams, 2007) confirmed normal productive phonology (M standard score = 110; $SD = 7$; range = 95 - 127), receptive vocabulary (M standard score = 112; $SD = 13$; range = 89 - 150), and expressive vocabulary (M standard score = 113; $SD = 11$; range = 95 - 135).

Stimuli. Stimuli are listed in the appendix with greater item-level detail provided in the supplemental materials (see Table S1). A pool of all legal English consonant-vowel-consonant (CVC) sequences was created (Storkel, In Press). This pool was submitted to an on-line calculator to identify real words in adult or child corpora (http://www.bncdnet.ku.edu/cml/info_ccc.vi), which were then eliminated from consideration as stimuli. In addition, only early acquired consonant sequences were retained (Storkel, In Press). This ensured that all remaining CVCs were nonwords with a high likelihood of correct production by preschool children ($n = 687$ CVCs). Two measures of probability and one measure of density were then calculated using the adult corpus. The adult corpus was selected because it was thought to reflect the language that children hear, thus providing a more accurate measure of children’s knowledge of probability and density; however, calculations based on either corpus are highly correlated (Storkel, In Press; Storkel & Hoover, 2010).

The two measures of phonotactic probability were positional segment sum and biphone sum (Storkel, 2004b). *Positional segment sum* was computed by summing the positional segment frequencies for each sound in the CVC. The positional segment frequency was computed by summing the log frequencies of the words in the corpus that contain the target sound in the target word position and then dividing by the sum of the log frequencies of the words in the corpus that contain any sound in the target word position. *Biphone sum* was computed in a similar manner except that the target is a pair of adjacent sounds rather than a single sound. *Density* was computed by counting the number of words in the corpus that differ from the target CVC by a one sound substitution, deletion, or addition in any word position (Storkel, 2004b).

Because only a limited number of nonwords can reasonably be taught to young children during an experimental study, the stimuli selection method needed to ensure that the trained items would adequately sample the full distribution of phonotactic probability values. To accomplish this, percentiles/quartiles were computed for the CVC pool and used to define a range of values for sampling different points of the phonotactic probability distribution. Specifically, lowest probability was defined as a positional segment sum and biphone sum below the 25th percentile (i.e., 1st quartile); midlow corresponded to the 25th to 49th percentile (i.e., 2nd quartile); midhigh corresponded to the 50th to 74th percentile (i.e., 3rd quartile); highest was the 75th percentile and above (i.e., 4th quartile). Five nonwords were then pseudo-randomly selected from each phonotactic probability quartile. Selection was pseudo-random because control of neighborhood density was considered as well as phonological similarity among the selected nonwords (i.e., an attempt was made to select dissimilar nonwords). Generally, the five items selected in a given phonotactic probability category sampled the full range of values in that quartile. That is, the sampled items approximated the minimum and maximum value that defined the quartile as well as included values between the minimum and maximum. However, the requirement to control neighborhood density (see next) somewhat truncated the items that could

be selected in the lowest and highest quartiles. Specifically, items below (approximately) the 10th percentile in the lowest category and items above the (approximately) 90th (segment sum) or 95th (biphone sum) percentile in the highest category could not be selected while controlling neighborhood density. Thus, the most extreme values at the beginning and end of the distribution of phonotactic probability are not well represented in the selected stimuli. Note that this approach to stimuli selection also has the added benefit of connecting the current stimuli to those used in prior studies, which have typically defined “low” and “high” probability using a median (i.e., 50th percentile) split. Thus, nonwords in the lowest and midlow probability quartiles in the current study generally correspond to “low” probability in past research; whereas, nonwords in the midhigh and highest quartiles correspond to “high” in past studies.

In terms of the control variable of neighborhood density, percentiles for the CVC pool were used to define acceptable values. Specifically, density was held constant at a mid-level, operationally defined as within 0.50 standard deviations of the 50th percentile. Table 1 shows the characteristics of the selected CVCs, with added detail shown in Table S1 of the supplemental materials.

For all analyses, a single measure of phonotactic probability was needed. Because positional segment sum and biphone sum are on different measurement scales, each value was converted to a z score based on the means and standard deviations of the stimulus pool (i.e., 687 CVC nonwords) and then averaged to yield one measure of phonotactic probability for analyses and figures. For analyses, this average z score was further re-scaled by multiplying by 10 to avoid extremely large or small odds ratios for the fixed effect of phonotactic probability, especially over the compressed range of lowest and midlow phonotactic probability.

Nonobjects were selected from a pool of 88 black and white line drawings developed by Kroll and Potter (1984) with additional normative data from Storkel and Adlof (2009). Twenty nonobjects were selected and paired with the twenty nonwords such that the objectlikeness

1 ratings (Kroll & Potter, 1984) and semantic set size (Storkel & Adlof, 2009) were matched
2
3
4
5 across the probability conditions (see appendix). In addition, the pairing of nonobjects to
6
7
8 nonwords was counterbalanced across participants.

9
10 *Procedures.* The twenty nonword-nonobject pairs were divided into five training sets of
11
12 four items, with each probability quartile represented in each set. That is, each training set
13
14 consisted of one lowest, one midlow, one midhigh, and one highest probability nonword (refer to
15
16 Table S1 of the supplemental materials for specific nonwords in each training set). Children were
17
18 trained on each of the sets on a different day using a different child-appropriate game context
19
20 (e.g., bingo, card game, board game). Training was administered via computer with
21
22 accompanying hard copy pictures of the nonobjects (e.g., bingo board, small cards, board game)
23
24 for game play. Training was divided into three blocks with each block providing eight auditory
25
26 exposures to each nonword-nonobject pair for a total of 24 cumulative auditory exposures.
27
28
29
30

31 Within a training block, presentation of nonword-nonobject pairs was randomized by the
32
33 computer. Each exposure consisted of the nonobject appearing centered on the computer screen
34
35 accompanied by a series of carrier phrases containing the corresponding nonword. The exact
36
37 exposure script was: “This is a *nonword*. Say *nonword*.” The child attempted to imitate the
38
39 nonword but no feedback was provided. “That’s the *nonword*. Remember, it’s a *nonword*. We’re
40
41 going to play a game. Find the *nonword*.” Here, the child would find the hard copy picture that
42
43 matched the picture on the computer screen and respond in a way appropriate to the game (e.g.,
44
45 move the marker on the game board to the corresponding picture). No feedback was provided.
46
47
48 “That’s the *nonword*. Say *nonword*.” Again, the child attempted to imitate the nonword but no
49
50 feedback was provided. “Don’t forget the *nonword*.” Thus, the training script provided eight
51
52 auditory exposures to the nonword, two imitation opportunities, and one picture matching
53
54 opportunity. Repetition accuracy ranged from 45%-100% with a mean accuracy of 87% ($SD =$
55
56 12%). Picture matching accuracy ranged from 80%-100% with a mean accuracy of 99% ($SD =$
57
58
59
60

4%). The intent of this set of training activities was to provide repeated exposure to the nonword-nonobject pair as well as relatively easy retrieval practice via repetition and picture matching prompts.

Learning was measured in a picture-naming test administered immediately upon completion of training and 1-week after training. Children had to correctly produce the entire CVC name of the picture to be credited with an accurate response. Picture naming was chosen because prior studies suggested stronger effects of word characteristics on expressive measures of word learning than on receptive measures of word learning (Storkel, 2001, 2003).

Analysis Approach. The data were analyzed using multilevel modeling. Multilevel modeling (MLM), also called mixed effects modeling, hierarchical linear modeling, or random coefficient modeling, is preferred over repeated measures ANOVA because it allows for a variety of variance/covariance structures, thus being more flexible regarding dependencies arising from repeated measures or missing and/or unbalanced data (Cnaan, Laird, & Slasor, 1997; Gueorguieva & Krystal, 2004; Hoffman & Rovine, 2007; Misangyi, LePine, Algina, & Goeddeke, 2006; Nezlek, Schroder-Abe, & Schutz, 2006; Quene & van den Bergh, 2004). Moreover, random effects of participants and items can be accommodated in the same analysis by incorporating crossed random intercepts, and this is becoming the favored analysis approach for psycholinguistic data (cf., Baayen, Davidson, & Bates, 2008; Locker, Hoffman, & Bovaird, 2007; Quene & van den Bergh, 2008). Note that the dependent variable for this study was accuracy (i.e., correct or incorrect), which is a binary variable. Thus, a logistic MLM was used.

The analysis proceeded in several steps. The first step was to examine the crossed random effects of participants and items to determine the significance and relative magnitude of participant and item (nonword) variance components in an empty model with no fixed predictors. For this particular experiment, the predictor phonotactic probability had a one-to-one relationship with nonword. That is, every nonword had a unique phonotactic probability so there

are no repeated items at a given phonotactic probability. Thus, the crossed random intercept for nonword is not needed in the subsequent models that include the fixed effect of phonotactic probability. For this reason, between-item variability not related to phonotactic probability is relegated to the residual variance component in this experiment, and this should be kept in mind when appraising the magnitude of fixed effects and between-subject variability.

The second step was to add the fixed effects of phonotactic probability, time (immediate vs. delayed test), and age (in months) to address the research questions. This model of the fixed effects used a spline regression model to capture the effect of phonotactic probability. Spline regression is a nonparametric approach used to approximate a nonlinear response across a continuous predictor without parametric assumptions or costs incurred by categorization (Marsh & Cormier, 2002). With linear splines, the effect of an explanatory variable (i.e. phonotactic probability) is assumed to be piecewise linear on a specified number of segments separated by knots (Gould, 1993; Panis, 1994). In terms of interpretation of the linear spline coefficients, coding can be for the slope in each segment or the change in slope from the prior segment. While the ability of linear spline models to provide a smooth transition across knots is generally valued, it is also possible to explore discontinuities between segments by dummy coding for an intercept change at each successive knot/segment (for an example of intercept dummy coding with linear splines, see UCLA Statistical Consulting Group). Note that dummy coding for change in level (as opposed to the actual level) in each segment is similar to ordinal dummy coding (Lyons, 1971). Many alternate codings for intercept and slope are possible. For this analysis, change in slope and intercept (level) from the prior segment was coded because the associated coefficients provide a test for whether a change/discontinuity is present without post-hoc tests. The number of segments is also arbitrary, but four segments were used in this analysis to align with the stimulus generation procedure based on quartiles. The specific coding scheme employed can be found in Table S3 of the online supplemental materials. Taken together, the fully-segmented

spline model allows for detection of discontinuity and nonlinearity in the relationship between phonotactic probability and word learning accuracy.

Although the fully-segmented spline model best matches the stimulus generation procedures, it is not the most parsimonious model. Thus, in a third and fourth step, alternative models were considered. Specifically, in the third step, phonotactic probability was modeled as a continuous linear predictor to determine whether the nonlinearity and discontinuity allowed by the spline model is really needed. Note that all other predictors in the linear model are the same as in the spline model, allowing for direct comparison between the two models using a likelihood-ratio test. In the fourth and final analysis step, phonotactic probability was modeled using a low-high median split for comparison to past studies of dichotomously coded phonotactic probability. Again, the other predictors in the model are the same as those in the spline model.

To facilitate insight into the magnitude of individual differences, participant level (and item level) variance was expressed as a median odds ratio (MOR, Merlo et al., 2006). Conceptually, the MOR conveys the median increase in the odds of a correct response between a pair of participants or items that are alike on all other covariates. Therefore, a MOR of 1 would indicate no change in the odds of a correct response as participants (or items) are changed. In complement, a large MOR would suggest substantial variability between participants (or items), indicating a large change in the odds of a correct response as participants (or items) are changed. The MOR has the further advantage of being on the same scale as the odds ratio (OR), which was used as the effect size for the fixed effects (e.g., phonotactic probability). In this way, the MOR for the random effects can be compared to the OR of the fixed effects to permit comparison of the magnitude of the effect of model predictors to the magnitude of individual differences (i.e., unexplained between-subject and between-item variances).

Results

Table 2 summarizes the different models created across analysis steps. The first model was an empty model with crossed random intercepts for both participants and items to assess baseline variability (see first column of Table 2). The MOR for participants was 2.34 (95% CI = 1.84-3.27) and for items was 1.48 (95% CI = 1.24-2.08). Thus, the variability between subjects is associated with a median difference of 2.34 in the odds of a correct response between two randomly drawn participants. Likewise, the variability between items is associated with a median difference of 1.48 in the odds of a correct response. However, recall that each item had a unique phonotactic probability z score, meaning that the item intercept term was dropped in all subsequent models, which included phonotactic probability. Thus, a second empty model with only a random intercept for participants was included. This is shown in the second column of Table 2. Note that the MOR (i.e., MOR = 2.31, 95% CI = 1.82-3.24) for participants is similar to the crossed random model that included items.

The next model was the fully-segmented spline model. As shown in the third column of Table 2, the spline model included a random intercept for participants, three fixed intercept parameters (each coding the change in level across segments), four slope parameters (each coding the change in slope from prior segment) as well as time (immediate vs. delayed test) and age (in months). The main effect of time was significant. The odds of a correct response were 2.23 (95% CI = 1.55-3.20) times lower in the delayed test than in the immediate test condition, indicating that significant forgetting occurred across this no-training gap. In terms of raw values, percent correct in the delayed test ($M = 5.76\%$, $SD = 8.99$) was lower than in the immediate test ($M = 11.14\%$, $SD = 9.66$). There also was a significant effect of age. Specifically, the odds of a correct response were 1.03 (95% CI = 1.00-1.06) times higher for a child one month older than another child. Note that the lower end of the confidence interval includes 1.00, which would normally indicate a non-significant effect. However, this is an artifact of rounding. In terms of

raw values, percent correct for the 5-year-olds ($M = 10.20\%$, $SD = 10.31$) was higher than for the 3-year-olds ($M = 6.63\%$, $SD = 6.51$). These effects can be seen in more detail in Figure S1 of the supplemental materials.

Turning to the main variable of interest, namely phonotactic probability, the top panel of Figure 1 aids visualization of the fully-segmented spline model. This panel shows the relationship between phonotactic probability on the x-axis and proportion correct on the y-axis collapsed across time and age. Vertical gray lines indicate the dividing points for the four probability quartiles: lowest, midlow, midhigh, highest. The four solid lines are a linear fit to each of the segments. These four lines closely approximate the splines that are modeled in the analysis. None of the intercepts in the spline model were significant. This indicates that the relationship between word learning accuracy and probability can be thought of as continuous. However, the slope for the first spline was significantly different from zero. As can be seen in Figure 1 and Table 2, the spline corresponding to lowest probability has a significant rising slope, indicating that the odds of a correct response were 1.69 (95% CI = 1.02-2.79) times higher for a one unit (i.e., 1/10 z score) increase in phonotactic probability in the lowest phonotactic probability quartile. The slope for the second spline also was significant. Here, the interpretation is that the slope for the second spline (midlow probability) is significantly different from the slope of the first spline (lowest probability). As shown in Figure 1, the second spline (midlow probability) has a falling slope, indicating that the odds of a correct response were 1.93 (95% CI = 1.15-3.23) times *lower* for a one unit increase in phonotactic probability in the midlow phonotactic probability quartile. The other two slope parameters (midhigh and highest probability) were not significant. The final two splines (midhigh and highest probability) are relatively flat, indicating minimal change in word learning accuracy as probability increased in these final two segments, which correspond to phonotactic probability above the median.

Turning to the alternative more parsimonious models, the effects of time and age in these models were similar to that of the fully-segmented spline model (see Table 2). Thus, the presentation of the alternative models focuses exclusively on the effect of phonotactic probability. The linear model is shown in the fourth column of Table 2. Recall that the difference between the linear and spline models is that phonotactic probability is now modeled with just one slope parameter (see dashed line in top panel of Figure 1). This forces the effect of phonotactic probability to be continuous and linear in this model, rather than allowing for discontinuity and nonlinearity, as in the fully-segmented spline model. The effect of phonotactic probability remained significant in the linear model with accuracy increasing as phonotactic probability decreased. However, the fully-segmented spline model provided significantly better fit to the data, $\chi^2(6) = 15.08, p = .02$. This indicates that the spline model better captures the effect of phonotactic probability.

Finally, the low-high median split model is shown in the last column of Table 2. Remember that this model was included to provide a comparison to past studies of phonotactic probability. Here, phonotactic probability is modeled with a second intercept term, capturing change in level across the median/50th percentile. Consistent with previous findings (Hoover, et al., 2010; Storkel, 2009; Storkel, et al., 2006; Storkel & Lee, 2011), participants were more accurate responding to low than high phonotactic probability items. However, once again, the fully-segmented spline analysis provided a better fit to the data, $\chi^2(6) = 13.71, p = .03$.

Taken together, word learning accuracy and probability showed a non-linear relationship that was not well captured by a simple linear slope across the entire distribution or a simple change in level (i.e., low vs. high) at the median of the distribution. Specifically, accuracy increased as probability increased in the lowest probability quartile. Then, accuracy decreased as probability increased in the midlow probability quartile. In the midhigh and highest probability quartile (i.e., above the median), accuracy was relatively stable and poor. Thus, there appeared to

be a change in the relationship between word learning accuracy and probability that occurred between the lowest and midlow probability quartiles, followed by no change above the median (i.e., midhigh and highest quartile). To investigate the location of the change point, a follow-up change point analysis was conducted (McArdle & Wang, 2008). The change point analysis estimates the location of the change rather than forcing the change in slope to occur between pre-defined segments corresponding to our probability quartiles (i.e., between lowest and midlow probability quartiles). The change-point analysis located the change point at phonotactic probability values near the minimum probability of the second spline (i.e., $z = -0.65$, see Table 1 for corresponding raw values). Thus, our somewhat arbitrarily chosen ranges seem to be capturing the location of the change-point, rather than biasing the location of the change point.

A final caveat relates to the variability across participants. There was significant variability across participants (i.e., Participant MOR = 2.37, 95% CI = 1.85-3.33). This participant variability was examined via (1) visual inspection of a figure plotting accuracy by phonotactic probability for individual participants (see Figure S2 of the supplemental materials), and (2) fitting several models with random coefficients for slopes (see Figure S3 of the supplemental materials). Based on these methods, variability appeared to be due to overall differences in accuracy rather than differences in the effect of phonotactic probability across participants. That is, some participants learned words with greater accuracy than other participants, which is captured by the random effect of participants, but all participants showed a roughly similar pattern in the effect of phonotactic probability on word learning, which is captured by the fixed effect of phonotactic probability.

Experiment 2: Neighborhood Density

Learning of novel words varying in neighborhood density (i.e., lowest, midlow, midhigh, highest) but matched in phonotactic probability and nonobject characteristics (i.e., objectlikeness ratings, number of semantic neighbors) was examined.

Method

Participants. Thirty-three 3-year-old ($M = 3$ years; 6 months; $SD = 0; 3$; range = 3; 1 – 3; 11) and 37 5-year-old ($M = 5$ years; 3 months; $SD = 0; 3$; range = 5; 0 – 6; 0) children meeting the same criteria as Experiment 1 participated. Children exhibited normal productive phonology (M standard score = 110; $SD = 9$; range = 83 - 127), receptive vocabulary (M standard score = 113; $SD = 12$; range = 88 - 147), and expressive vocabulary (M standard score = 113; $SD = 10$; range = 93 - 135). None of the children participated in Experiment 1.

Stimuli. Stimuli are shown in the appendix with more detailed item data in Table S2 of the on-line supplemental materials. Nonword stimuli were selected following the procedures outlined for Experiment 1, except that neighborhood density was the independent variable and the two measures of phonotactic probability were controlled. As with Experiment 1, the approach to stimuli selection led to adequate sampling of neighborhood density values from approximately the 10th to the 95th percentile, but extreme values were not sampled due to the need to control phonotactic probability (see Table 1). The same nonobjects used in Experiment 1 were used here.

Procedures. Procedures were identical to Experiment 1. In terms of responses during training, repetition accuracy ranged from 60%-100% with a mean accuracy of 89% ($SD = 9\%$). Picture matching accuracy ranged from 44%-100% with a mean accuracy of 98% ($SD = 8\%$).

Analysis Approach. Analysis approach was similar to Experiment 1 with the exception of the use of z scores. Because there was only one measure of neighborhood density, raw values were used in the analyses and figure rather than z scores. A second difference from Experiment 1 is that several nonwords had the same density, making it possible to disentangle neighborhood density and between-item variability. Thus, the crossed random effects of participants and items are included in all models. Table S4 of the supplemental materials provides the model coding.

Results

Table 3 summarizes the four models. Beginning with the empty model in the first column of Table 3, participants and items were modeled as crossed random effects. The MOR for participants was 2.31 (CI = 1.89-3.00) and the MOR for items was 2.26 (CI = 1.73-3.36). Thus, the variability between participants is associated with a median difference of 2.31 in the odds of a correct response between two randomly drawn participants. Likewise, the variability between items is associated with a median difference of 2.26 in the odds of a correct response between randomly drawn items. Recall that there were several items with the same neighborhood density. Thus, unlike Experiment 1, the random intercept for items is retained in all subsequent models.

Turning to the fully-segmented spline model in the second column of Table 3, fixed effects were added to the empty model. As in Experiment 1, neighborhood density was modeled with three intercept change terms and four slope terms. Effects of time (immediate vs. delayed test) and age (in months) also were included. Once again, there was a significant effect of time. The odds of a correct response were 1.97 (95% CI = 1.47-2.65) times lower in the delayed test than in the immediate test. In terms of raw values, percent correct in the delayed test ($M = 6.46\%$, $SD = 7.97$) was lower than in the immediate test ($M = 11.11\%$, $SD = 10.36$). Likewise, the effect of age was significant. Specifically, the odds of a correct response were 1.03 (95% CI = 1.00-1.05) times higher for a child one month older than another child. In terms of raw values, percent correct for the 5-year-olds ($M = 10.46\%$, $SD = 10.24$) was higher than for the 3-year-olds ($M = 6.95\%$, $SD = 5.61$). These effects can be seen in more detail in Figure S4 of the supplemental materials.

More important, however, is the influence of neighborhood density. The lower panel of Figure 1 provides a visualization of the model, by showing the relationship between density and proportion correct collapsed across time and age. Again, vertical gray lines indicate the dividing points for the four density quartiles: lowest, midlow, midhigh, highest. The four solid lines are

1
2 the linear fit to each of the density segments. The results were very straightforward. None of the
3
4 intercept terms were significant, indicating no discontinuity between segments. In addition, none
5
6 of the slope terms were significant, indicating that the distribution may be best described by a
7
8 single slope. This possibility is explored in the alternative models. As shown in the bottom panel
9
10 of Figure 1, word learning accuracy appears to increase as neighborhood density increases.
11
12

13
14 The alternative models did examine the effects of time and age, with no major differences
15
16 in findings from the spline model. These effects can be seen in more detail for the linear model
17
18 in Figure S4 of the supplemental materials. Presentation of the alternative models focuses solely
19
20 on the effect of neighborhood density. The third column of Table 3 shows the linear model.
21
22 Recall that this model uses a single slope parameter to capture the effect of neighborhood
23
24 density, making it more parsimonious than the spline model. The effect of density was
25
26 significant. Specifically, the odds of a correct response were 1.09 (95% CI = 1.02-1.16) times
27
28 higher for a one neighbor increase in density across the full distribution of density values.
29
30 Importantly, there was no difference in fit between the spline model and this linear model, $\chi^2(6)$
31
32 = 5.79, $p = .45$. This suggests that the more parsimonious linear model should be preferred over
33
34 the fully-segmented spline model. Thus, the relationship between word learning and density is
35
36 best described as a continuous linear function. Note that the MOR for items in this model is 2.00
37
38 (95% CI = 1.58-2.85), which can be directly compared to the OR for neighborhood density,
39
40 which is 1.09 (95%CI = 1.02-1.16). From this comparison, it is clear that neighborhood density
41
42 is not the only item characteristic that influences ease of word learning.
43
44
45
46
47
48
49

50 The final column of Table 3 reports the results of the median split model. Although the
51
52 comparison between low and high density did not reach significance ($p = .20$), the trend (i.e.,
53
54 better accuracy for high than low density) is in the same direction as past studies (Hoover, et al.,
55
56 2010; Storkel, 2009; Storkel, et al., 2006; Storkel & Lee, 2011). Again, there was no difference
57
58 in fit between the spline model and this median-split model, $\chi^2(6) = 10.18$, $p = .12$.
59
60

As with Experiment 1, participant variability was explored for the preferred model, namely the linear model. See Figure S5 in the supplemental materials. Again, participant variability appeared to be captured by a difference in overall accuracy rather than differences in the effect of density across participants. That is, some participants learned words with greater accuracy than other participants, which is captured by the random effect of participants, but all participants showed a roughly similar linear pattern in the effect of neighborhood density on word learning.

General Discussion

The goal of the present investigation was to determine whether incremental changes in phonotactic probability and neighborhood density influenced word learning performance and, if so, to determine the precise pattern of the relationship between probability or density and word learning. Both studies showed that incremental changes in phonotactic probability and neighborhood density did influence word learning. Moreover, the pattern of word learning performance across the phonotactic probability distribution differed from the pattern of word learning performance across the neighborhood density distribution. For phonotactic probability, a nonlinear pattern was observed. Specifically, word learning improved as probability increased in the lowest probability quartile. Then, there was a change in the next quartile (i.e., midlow probability) with word learning worsening as probability increased. In the midhigh and highest probability quartiles, word learning was relatively stable and poor. In contrast, word learning tended to improve as neighborhood density increased in a predominately linear fashion across the full density distribution. This finding of different patterns of word learning performance across the phonotactic probability distribution versus across the neighborhood density distribution partially supports the initial hypothesis that phonotactic probability and neighborhood density influence different word learning processes.

Phonotactic probability was hypothesized to influence triggering, namely allocation of a new representation. Based on past studies of disambiguation, word learning was expected to worsen as probability increased in the lower end of the phonotactic probability distribution and then expected to remain stable (and poor) at the higher end of the distribution. This hypothesis was partially supported, with the predicted pattern being observed in the midlow, midhigh, and highest probability quartiles. However, the finding that word learning improved as phonotactic probability increased in the lowest probability quartile was unexpected and appears inconsistent with claims about the triggering process. Therefore, the role of phonotactic probability in word learning may need to be reconsidered. One possibility is that phonotactic probability is involved in two aspects of word learning: recognizing which sound sequences are potential words and recognizing which sound sequences are novel words to-be-learned (i.e., triggering). In fact, past studies suggest that infants do not accept every sound, even every sequence of speech sounds, as a potential or acceptable word (Balaban & Waxman, 1997; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007; MacKenzie, Curtin, & Graham, 2012). It is possible that phonotactic probability could influence recognition of which sound sequences are acceptable words, although this hypothesis awaits empirical testing. The implication for word learning is that children would not learn sound sequences that fail to meet some sort of acceptability criteria for their language.

The tentative account of the current findings is that in the lowest phonotactic probability quartile, the sound sequences are unusual for the language. In support of this, in a sample of 1,396 CVC real words (Storkel, In Press), only 3% of the sample had positional segment sums and biphone sums in the same range as the nonwords in our lowest phonotactic probability quartile. Within this lowest probability quartile, recognition that the sound sequence is an acceptable or potential word in the language may increase as probability increases, potentially accounting for the observed pattern in Figure 1. Presumably a threshold is crossed at the juncture between lowest and midlow probability, and all sound sequences with higher probability are

1
2
3 recognized as acceptable words. Note that 10% of the 1,396 CVC real words had positional
4
5 segment sums and biphone sums in the same range as the nonwords in our midlow phonotactic
6
7 probability quartile, confirming that these sound sequences were not as unusual in the language.
8
9 At this point (i.e., midlow probability), the triggering role of probability becomes more visible,
10
11 such that recognition that a sound sequence is a novel word, requiring learning, decreases as
12
13 probability increases. Then, at the median, performance stabilizes at a low level of word learning
14
15 accuracy. These midhigh and highest probability sound sequences are likely recognized as
16
17 acceptable words in the language but are not particularly novel based on their sound sequence
18
19 alone. It is likely that other characteristics, many of which were controlled in the current
20
21 research, would be more influential in triggering learning for these sound sequences and their
22
23 referents. Taken together, the modified account is that word learning only occurs for sound
24
25 sequences that are acceptable and novel, with phonotactic probability contributing to both
26
27 criteria.
28
29
30
31
32

33 The finding of a linear relationship between neighborhood density and word learning is
34
35 consistent with the hypothesis that density influences configuration. Specifically, working
36
37 memory is argued to affect configuration by providing temporary storage of the sound sequence
38
39 while the new representation is being created (Gupta & MacWhinney, 1997). When a sound
40
41 sequence is heard, existing lexical representations in long-term memory are activated. These
42
43 existing representations provide support to working memory such that the more representations
44
45 that are activated (i.e., the higher the density), the better the maintenance of a sound sequence in
46
47 working memory (Roodenrys & Hinton, 2002; Thomson, Richardson, & Goswami, 2005; Thorn
48
49 & Frankish, 2005). A related point is that existing representations may be more detailed (i.e.,
50
51 segmental) when there are many similar representations (i.e., the higher the density, Garlock,
52
53 Walley, & Metsala, 2001; Metsala, Stavrinos, & Walley, 2009; Metsala & Walley, 1998; Storkel,
54
55 2002). More detailed representations could lead to better maintenance of a sound sequence in
56
57
58
59
60

1
2
3 working memory (Metsala, et al., 2009). For configuration, better maintenance of a sound
4
5 sequence in working memory translates into greater support for the creation of a complete and
6
7 accurate new lexical representation in long-term memory. Thus, as the number of existing lexical
8
9 representations activated increases or as the segmental detail of existing representations
10
11 increases, the quality or robustness of the new lexical representation likely increases.
12
13

14 Turning to the engagement process, recall that past research suggests that engagement
15
16 occurs late in word learning, resulting from memory consolidation processes during sleep
17
18 (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Leach & Samuel, 2007). Thus, the primary
19
20 evidence for engagement comes from changes in responding that occur over a delay interval
21
22 without further training. The current data are inconsistent with an explanation that appeals to
23
24 engagement because it appears that engagement may not have occurred. That is, performance in
25
26 both experiments significantly declined over the delay, suggesting an absence of engagement
27
28 (Dumay & Gaskell, 2007). Previous research suggests that participants sometimes encapsulate
29
30 words learned in the laboratory from the rest of the lexicon (Magnuson, Tanenhaus, Aslin, &
31
32 Dahan, 2003). This could account for the apparent lack of engagement in the current studies.
33
34
35
36

37 38 *Conclusion*

39
40 Past studies have examined only gross distinctions between low and high probability or
41
42 density. The current studies provide evidence that incremental changes in probability and density
43
44 influence word learning. Moreover, the pattern of word learning performance across the
45
46 phonotactic probability distribution differed from the pattern of word learning performance
47
48 across the neighborhood density distribution, supporting the theory that these two variables
49
50 influence different word learning processes. Specifically, phonotactic probability appeared to
51
52 influence two aspects of triggering word learning: (1) recognition of sound sequences as
53
54 acceptable words in the language; (2) recognition of sound sequences as novel to the child. In
55
56 contrast, neighborhood density seemed to influence configuration of a new representation in the
57
58
59
60

mental lexicon. Further examination of incremental changes in probability or density may yield new insights into other cognitive processes, such as spoken word recognition, learning, and memory.

Author Note

The project described was supported by Grants DC 08095, DC 05803, and HD02528 from NIH. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The second author was supported by the Analytic Techniques and Technology (ATT) Core of the Center for Biobehavioral Neurosciences of Communication Disorders (BNCD, DC05803). We would like to thank staff of the Participant Recruitment and Management Core (PARC) of the Center for Biobehavioral Neurosciences of Communication Disorders (BNCD, DC05803) for assistance with recruitment of preschools and children; staff of the Word and Sound Learning Lab (supported by DC 08095) for their contributions to stimulus creation, data collection, data processing, and reliability calculations; and the preschools, parents, and children who participated.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, 64(1), 3-26.

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16, 2349-2380.

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. *Developmental Science*, 1-18.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18(1), 35-39.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test—4th edition*. Circle Pines, MN: American Guidance Service.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy*, 4, 349-369.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, 105(1), 218-228.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, 45(3), 468-492.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89(2), 105-132.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513-543.

- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation-2*. Circles Pines, MN: American Guidance Service.
- Gould, W. (1993). Linear splines and piecewise linear functions. *Stata Technical Bulletin*, 15, 13-17.
- Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61, 310-317.
- Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language. Special Issue: Computer models of impaired language*, 59(2), 267-333.
- Hoffman, L., & Rovine, M. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101-117.
- Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language*, 63, 100-116.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning & Verbal Behavior*, 23(1), 39-66.
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, 55, 306-353.
- Li, P., Farkas, I., & Mac Whinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*(17), 1345-1362.
- Locker, L., Jr., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, 39(4), 723-730.

1
2
3 Lyons, M. (1971). Techniques for using ordinal measures in regression and path analysis.
4
5 *Sociological Methodology*, 3, 147-171.
6
7 MacKenzie, H., Curtin, S., & Graham, S. A. (2012). 12-month-olds' phonotactic knowledge
8
9 guides their word-object mappings. *Child Development*, 83, 1129-1136.
10
11 Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of
12
13 spoken word learning and recognition: Studies with artificial lexicons. *Journal of*
14
15 *Experimental Psychology: General*, 132(2), 202-227.
16
17
18 Marsh, L. C., & Cormier, D. R. E. S. R. M. V. S. P. (Eds.). (2002). *Spline Regression Models*
19
20 (Vol. 137). Thousand Oaks, CA: SAGE Publications.
21
22
23 McArdle, J. J., & Wang, L. (2008). Modeling age-based turning points in longitudinal life-span
24
25 growth curves of cognition. In P. Cohen (Ed.), *Applied data analytic techniques for*
26
27 *turning points research* (pp. 105-128). New York: Taylor & Francis.
28
29
30 Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006). A brief
31
32 conceptual tutorial of multilevel analysis in social epidemiology: using measures of
33
34 clustering in multilevel logistic regression to investigate contextual phenomena. *Journal*
35
36 *of epidemiology and community health*, 60, 290-297.
37
38
39
40 Metsala, J. L., Stavrinos, D., & Walley, A. C. (2009). Children's spoken word recognition and
41
42 contributions to phonological awareness and nonword repetition: A 1-year follow-up.
43
44 *Applied Psycholinguistics*, 30(1), 101-121.
45
46
47 Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental
48
49 restructuring of lexical representations: Precursors to phonemic awareness and early
50
51 reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning*
52
53 *literacy* (pp. 89-120). Hillsdale, NJ: Erlbaum.
54
55
56 Misangyi, V. F., LePine, J. A., Algina, J., & Goeddeke, F. (2006). The adequacy of repeated
57
58 measures regression for multilevel research: Comparisons with repeated measures
59
60

- ANOVA, multivariate repeated measures ANOVA, and multilevel modeling across various multilevel research designs. *Organizational Research Methods*, 9, 5-28.
- Nezlek, J. B., Schroder-Abe, M., & Schutz, A. (2006). Multilevel analyses in psychological research. Advantages and potential of multi-level random coefficient modeling. *Psychologische Rundschau*, 57, 213-223.
- Panis, C. (1994). Linear splines and piecewise linear functions. *Stata Technical Bulletin*, 18, 27-29.
- Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43, 103-121.
- Quene, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of memory & language*, 59(4), 413-425.
- Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 29-33.
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 1019-1034.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6), 1321-1337.
- Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, 29(2), 251-274.
- Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, 46(6), 1312-1323.
- Storkel, H. L. (2004a). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research*, 47(5), 1194-1212.

1
2
3 Storkel, H. L. (2004b). Methods for minimizing the confounding effects of word length in the
4
5 analysis of phonotactic probability and neighborhood density. *Journal of Speech,*
6
7 *Language, and Hearing Research, 47*(6), 1454-1468.
8
9
10 Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and
11
12 semantic variables on word learning by infants. *Journal of Child Language, 36*(2), 291-
13
14 321.
15
16
17 Storkel, H. L. (In Press). A corpus of consonant-vowel-consonant (CVC) real words and
18
19 nonwords: Comparison of phonotactic probability, neighborhood density, and consonant
20
21 age-of-acquisition. *Behavior Research Methods.*
22
23
24 Storkel, H. L., & Adlof, S. M. (2009). Adult and child semantic neighbors of the Kroll and Potter
25
26 (1984) nonobjects. *Journal of Speech, Language and Hearing Research, 52*(2), 289-305.
27
28
29 Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and
30
31 neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing*
32
33 *Research, 49*(6), 1175-1192.
34
35
36 Storkel, H. L., & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability
37
38 and neighborhood density based on child corpora of spoken American English. *Behavior*
39
40 *Research Methods, 42*, 497-506.
41
42
43 Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and
44
45 neighborhood density on lexical acquisition by preschool children. *Language and*
46
47 *Cognitive Processes, 26*, 191-211.
48
49
50 Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The
51
52 role of phonotactic probability and word frequency. *Journal of Child Language, 32*(4),
53
54 827-853.
55
56
57
58
59
60

- Thomson, J. M., Richardson, U., & Goswami, U. (2005). Phonological similarity neighborhoods and children's short-term memory: typical development and dyslexia. *Memory & Cognition*, 33(7), 1210-1219.
- Thorn, A. S., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 729-735.
- UCLA Statistical Consulting Group. How Can I Run a Piecewise Regression in Stata? Retrieved November 28, 2012, from <http://www.ats.ucla.edu/stat/stata/faq/piecewise.htm>
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325-329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1), 114-132.
- Williams, K. T. (2007). *Expressive vocabulary test-2nd edition*. Circle Pines, MN: American Guidance Services, Inc.

Appendix: Stimuli

Nonwords used in Experiment 1 and 2.

Experiment 1: Probability	Experiment 2: Density
Low – tɔf, huf, geɪg, haʊd, bug	Low – baf, jɪb, maf, paɪb, gɛp
Midlow – baɪb, hɔd, doʊb, gid, goʊm	Midlow – toʊb, doʊb, jun, waʊn, fɛg
Midhigh – poug, peɪb, fɛg, tab, moum	Midhigh – gut, wout, daɪp, hɛg, maɪp
High – pag, bɪf, poum, mɛm, dɪf	High – tip, beɪm, flɪm, mɪp, gaɪt

Nonobjects (Kroll & Potter, 1984) used in Experiment 1 and 2.

- Group 1 – nonobjects 11, 29, 38, 81, 86
- Group 2 – nonobjects 26, 27, 46, 59, 63
- Group 3 – nonobjects 31, 37, 67, 78, 80
- Group 4 – nonobjects 5, 22, 23, 53, 82

Table A1. *Nonobject characteristics by group.*

		Group 1	Group 2	Group 3	Group 4
Objectlikeness Rating	<i>M</i>	4.2	4.2	4.2	4.2
(Kroll & Potter, 1984)	<i>(SD)</i>	0.9	0.8	0.8	0.9
	range	3.3-5.2	3.2-5.2	3.4-5.2	3.4-5.6
Number of Semantic Neighbors	<i>M</i>	10.8	10.4	10.4	10.6
(Storkel & Adlof, 2009)	<i>(SD)</i>	0.8	0.5	1.1	0.9
	range	10-12	10-11	9-12	10-12
Pairing of nonobject groups to the nonword conditions (low, midlow, midhigh, high) was counterbalanced across participants.					

Table 1. *Characteristics of the nonwords in each experiment.*

Condition		Lowest	Midlow	Midhigh	Highest
Percentile		< 25 th	25 th – 49 th	50 th – 74 th	>= 75 th
Experiment 1: Probability					
Positional Segment Sum ¹	<i>M</i>	0.083	0.111	0.143	0.172
	<i>(SD)</i>	(0.007)	(0.016)	(0.010)	(0.009)
	range	0.073 - 0.091	0.094 - 0.127	0.131 - 0.156	0.163 - 0.183
Biphone Sum ¹		0.0015	0.0028	0.0049	0.0105
		(0.0001)	(0.0008)	(0.0010)	(0.0048)
		0.0014 - 0.0015	0.0018 - 0.0037	0.0041 - 0.0064	0.0070 - 0.0187
Density ¹		10	11	11	11
		(1)	(1)	(2)	(1)
		9 – 12	9 - 13	9 - 14	10 – 13
Experiment 2: Density					
Positional Segment Sum ¹	<i>M</i>	0.136	0.127	0.126	0.126
	<i>(SD)</i>	(0.006)	(0.006)	(0.008)	(0.008)
	range	0.130 - 0.145	0.120 - 0.137	0.114 - 0.136	0.113 - 0.135
Biphone Sum ¹		0.0039	0.0042	0.0047	0.0039
		(0.0016)	(0.0012)	(0.0012)	(0.0013)
		0.0026 - 0.0066	0.0027 - 0.0059	0.0031 - 0.0058	0.0027 - 0.0061
Density ¹		5	10	15	20
		(1)	(1)	(2)	(2)
		4 – 5	8 - 11	13 - 17	18 - 24

¹Based on the adult corpus and on-line calculator described in Storkel and Hoover (2010).

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

Variable	Crossed Random Empty Model	Participant Only Empty Model	Fully-segmented Spline Model	Linear Model	Median Split Model
Phonotactic Probability Intercept 2 ³			1.12 (0.41-3.04) ¹		1.76 (1.23-2.5)** ^{1 2}
Phonotactic Probability Intercept 3			1.01 (0.43-2.38) ¹		
Phonotactic Probability Intercept 4			1.66 (0.45-6.15)		
Phonotactic Probability Slope 1			1.69 (1.02-2.79)*	1.03 (1.01-1.05)** ¹	
Phonotactic Probability Slope 2			1.93 (1.15-3.23)* ¹		
Phonotactic Probability Slope 3			1.08 (0.87-1.34)		
Phonotactic Probability Slope 4			1.01 (0.83-1.22)		
Time			2.23 (1.55-3.20)*** ¹	2.20 (1.54-3.16)*** ¹	2.20 (1.54-3.16)*** ¹
Age			1.03 (1.00-1.06)*	1.03 (1.01-1.06)*	1.03 (1.01-1.06)*
MOR for Participants	2.34	2.31	2.28	2.24	2.24
(confidence interval)	(1.84-3.27)	(1.82-3.24)	(1.80-3.19)	(1.77-3.12)	(1.77-3.13)
MOR for Items	1.48	Not in Model	Not in Model	Not in Model	Not in Model
(confidence interval)	(1.24-2.08)				
Log-likelihood	-502.0	-505.4	-481.8	-489.4	-488.7

* p < .05, ** p < .01, *** p < .001

¹The reciprocal was taken for OR < 1. For these effects, the OR indicates that that a correct response is *less likely* for higher than a lower value of the variable. For all other ORs, the interpretation is that a correct response is *more likely* for a higher than lower value of the variable.

²Note that Phonotactic Probability Intercept 2 in this model is located at the median (i.e., start of the midhigh quartile).

³All models included an Intercept 1 term that serves as the traditional constant, and is the OR denominator.

Table 3. *Models from Experiment 2. Estimates are expressed as odds ratios with 95% confidence interval in parentheses.*

Variable	Crossed Random Empty Model	Fully-segmented Spline Model	Linear Model	Median Split Model
Neighborhood Density Intercept 2 ³		5.49 (0.04-863) ¹		1.69 (.76-3.76) ²
Neighborhood Density Intercept 3		5.00 (0.72-35) ¹		
Neighborhood Density Intercept 4		1.41 (0.26-7.56)		
Neighborhood Density Slope 1		1.95 (0.46-8.24) ¹	1.09 (1.02-1.16)**	
Neighborhood Density Slope 2		1.48 (0.32-6.89) ¹		
Neighborhood Density Slope 3		1.07 (0.53-2.17) ¹		
Neighborhood Density Slope 4		1.29 (0.76-2.21) ¹		
Time		1.97 (1.47-2.65)*** ¹	1.97 (1.47-2.65)*** ¹	1.97 (1.47-2.65)*** ¹
Age		1.03 (1.00-1.05)*	1.03 (1.00-1.05)*	1.03 (1.00-1.05)*
MOR for Participants	2.31	2.25	2.25	2.25
(confidence interval)	(1.89-3.00)	(1.85-2.91)	(1.85-2.92)	(1.86-2.92)
MOR for Items	2.26	1.77	2.00	2.20
(confidence interval)	(1.73-3.36)	(1.44-2.45)	(1.58-2.85)	(1.69-3.24)
Log-likelihood	-727.8	-709.3	-712.2	-714.4

* p < .05, ** p < .01, *** p < .001

¹The reciprocal was taken for OR < 1. For these effects, the OR indicates that that a correct response is *less likely* for higher than a lower value of the variable. For all other ORs, the interpretation is that a correct response is *more likely* for a higher than lower value of the variable.

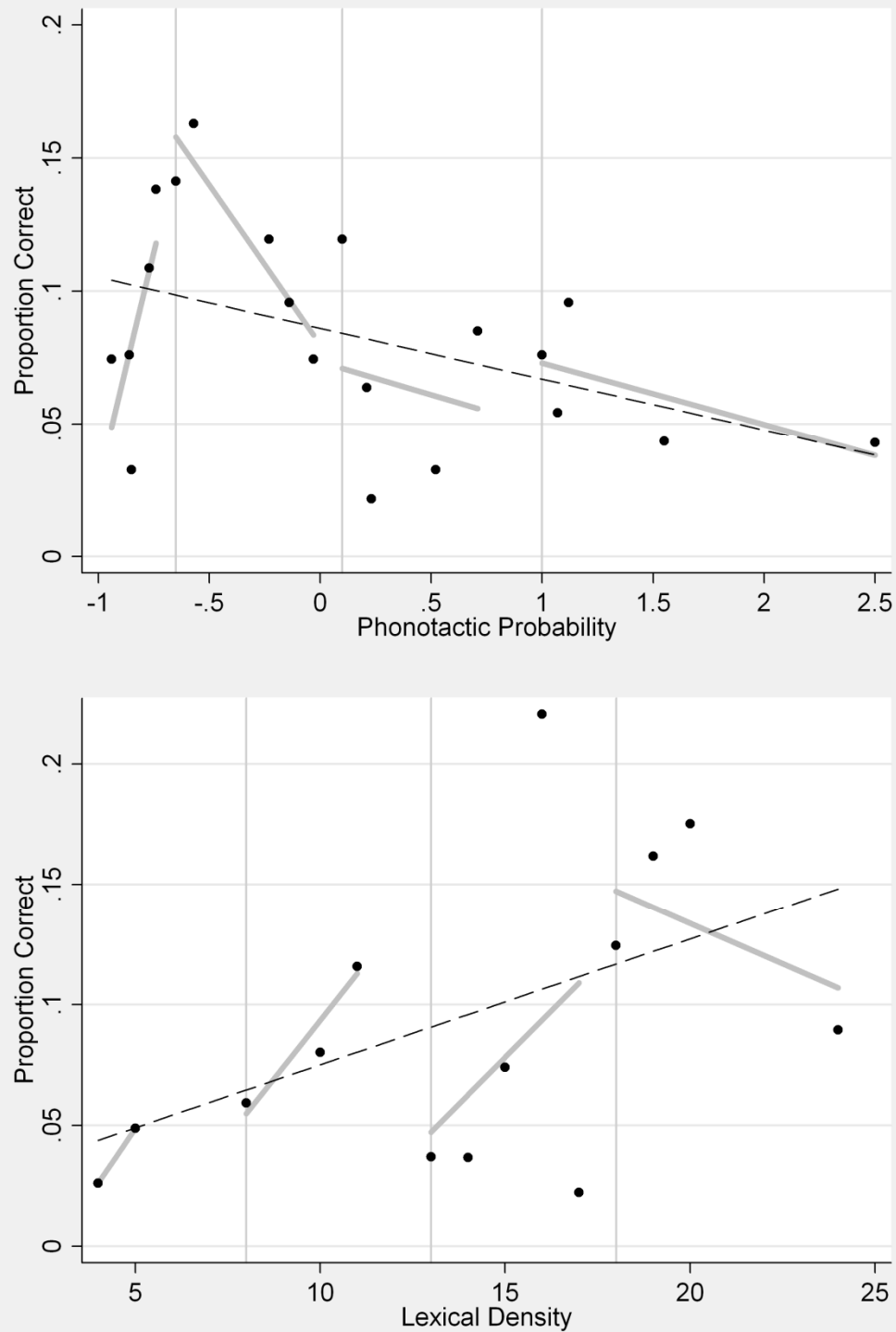
²Note that Neighborhood Density Intercept 2 in this model is located at the median (i.e., start of the midhigh quartile).

³All models included an Intercept 1 term that serves as the traditional constant, and is the OR denominator.

Figure Caption

Figure 1. Mean proportion correct for Experiment 1: Phonotactic Probability (top panel, z scores) and Experiment 2: Neighborhood Density (bottom panel, raw values), collapsed across time and age. Circles represent individual nonwords (Experiment 1: Phonotactic Probability) or mean proportion correct across nonwords with the same neighborhood density (Experiment 2: Neighborhood Density). Vertical grey lines indicate the dividing points between the four quartiles of the distribution: lowest, midlow, midhigh, highest. The four solid lines are the linear fit lines for each quartile (i.e., lowest, midlow, midhigh, highest). This corresponds to the fully-segmented spline model. The dashed line is the linear fit line for the full distribution. This corresponds to the linear model.

Figure 1



Supplementary Materials
Summary Description for Web Portal

This pdf document contains:

- Two tables that provide detailed information about the nonwords used in Experiment 1: Phonotactic Probability (Table S1) and Experiment 2: Neighborhood Density (Table S2). Available variables include: positional segment sum, biphone sum, neighborhood density, and training set.
- Two tables providing the coding scheme used for the fully-segmented spline model in Experiment 1: Phonotactic Probability (Table S3) and Experiment 2: Neighborhood Density (Table S4). Additional tutorial is provided regarding alternative coding schemes (see Table S3) for spline regression.
- Two figures providing more detailed visualization of the effects of Time (immediate vs. delayed), Age, and Experiment 1: Phonotactic Probability (Figure S1) or Experiment 2: Neighborhood Density (Figure S4).
- Three figures showing individual variation in responding across participants for Experiment 1: Phonotactic Probability (Figures S2 and S3) and Experiment 2: Neighborhood Density (Figure S5).

Supplemental Materials

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S. Y. (In Review). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech-Language-Hearing Research*.

Table S1: Nonword Stimuli for Experiment 1

Nonword	Phonotactic Probability Quartile	SegSum	ZSegSum	BiphSum	ZBiphSum	PhonProb Avg Z-Score	Density	Training Set
geɪg	Lowest	0.07	-1.19	0.002	-0.69	-0.94	10	C
huf	Lowest	0.08	-1.00	0.001	-0.72	-0.86	10	B
tɔf	Lowest	0.08	-1.01	0.002	-0.69	-0.85	9	A
haʊd	Lowest	0.09	-0.86	0.001	-0.69	-0.77	12	D
bug	Lowest	0.09	-0.76	0.001	-0.72	-0.74	10	E
hɔd	Midlow	0.09	-0.69	0.002	-0.61	-0.65	10	B
gid	Midlow	0.10	-0.65	0.002	-0.48	-0.56	13	D
baɪb	Midlow	0.11	-0.27	0.003	-0.18	-0.23	9	A
doub	Midlow	0.13	0.10	0.003	-0.37	-0.13	11	C
goum	Midlow	0.12	0.05	0.004	-0.10	-0.03	11	E
tab	Midhigh	0.13	0.20	0.004	0.00	0.10	13	D
fɛg	Midhigh	0.14	0.35	0.004	0.06	0.20	11	C
peɪb	Midhigh	0.14	0.40	0.004	0.06	0.23	10	B
poug	Midhigh	0.15	0.69	0.005	0.35	0.52	9	A
moum	Midhigh	0.16	0.79	0.006	0.62	0.71	14	E
bɪf	Highest	0.17	1.06	0.008	0.94	1.00	11	B
pɑg	Highest	0.16	0.96	0.008	1.18	1.07	10	A
poum	Highest	0.18	1.45	0.007	0.78	1.11	13	C
mɛm	Highest	0.18	1.36	0.011	1.74	1.55	12	D
dɪf	Highest	0.17	1.08	0.019	3.91	2.49	11	E

Table S1.Note 1: This table shows the item level data for each nonword in Experiment 1. From left to right, the table contains (1) IPA transcription of each nonword; (2) phonotactic probability quartile; (3) raw positional segment sum; (4) positional segment sum converted to z score; (5) raw biphone sum; (6) biphone sum converted to z score; (6) average of the segment sum and biphone sum z scores (used in all analyses); (7) raw neighborhood density; (8) assigned training set (e.g., set A trained on one day; set B on another; etc.).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table S2: Nonword Stimuli for Experiment 2					
Nonword	Density Quartile	SegSum	BiphSum	Density	Training Set
jɪb	Lowest	0.13	0.003	4	B
gɛp	Lowest	0.14	0.003	4	E
baf	Lowest	0.13	0.004	5	A
maf	Lowest	0.14	0.007	5	C
parb	Lowest	0.14	0.003	5	D
waʊn	Midlow	0.13	0.004	8	D
toub	Midlow	0.12	0.004	10	A
doub	Midlow	0.13	0.003	11	B
jun	Midlow	0.13	0.006	11	C
fɛg	Midlow	0.14	0.004	11	E
hɛg	Midhigh	0.13	0.006	13	D
maɪp	Midhigh	0.13	0.004	14	E
gut	Midhigh	0.11	0.003	15	A
daɪp	Midhigh	0.12	0.005	16	C
wout	Midhigh	0.14	0.006	17	B
mɪp	Highest	0.13	0.004	18	D
gaɪt	Highest	0.13	0.004	18	E
fʌm	Highest	0.14	0.006	19	C
tip	Highest	0.11	0.003	20	A
berm	Highest	0.13	0.003	24	B

Table S2.Note 1: This table shows the item level data for each nonword in Experiment 2. From left to right, the table contains (1) IPA transcription of each nonword; (2) neighborhood density quartile; (3) raw positional segment sum; (4) raw biphone sum; (5) raw neighborhood density (used in all analyses); (6) assigned training set (e.g., set A trained on one day; set B on another; etc.).

Spline regression is a nonparametric approach used to approximate nonlinear response across a continuous predictor without parametric assumptions/artifacts, or costs incurred with categorization (Marsh & Cormier, 2002). With linear splines the effect of the explanatory variable is assumed to be piecewise linear on some number of segments demarcated by knots (Gould, 1993; Panis, 1994). Linear spline coefficients are straightforward to interpret. Coding can be for slope in each segment (i.e., absolute coding scheme in blue in Table S3), or change in slope from the previous segment (i.e., change coding scheme in green in Table S3). While the ability of linear spline models to provide a smooth continuous transition across knots is generally valued, it is also possible to explore discontinuities by additionally dummy coding for an intercept change at each successive knot/segment (for an example of intercept dummy coding with linear splines, see UCLA Statistical Consulting Group). Dummy coding for change in level (as opposed to absolute level) in each segment is similar to ordinal dummy coding (Lyons, 1971). Many alternate codings for level and slope are possible, and we have chosen to code for change in slope and change in level from prior segment because the associated coefficients provide a test for whether a change/discontinuity is present without post-hoc tests (i.e., change coding scheme in green in Table S3 and S4). The number of segments is also arbitrary, but we employ four segments to align with our stimulus generation based on quartiles.

Table S3: Segment Coding for Experiment 1

Predictor		seg	Slope Coding								Level Coding								Freq
			Absolute Segment Slope				Segment Change in Slope				Absolute Segment Level				Segment Change in Level				
											i1	i2	i3	i4					
PhonProb	x10		s1	s2	s3	s4	cs1	cs2	cs3	cs4	i1	i2	i3	i4	i1x	i2x	i3x	i4x	
-0.94	-9.4	1	-9.4	0	0	0	-9.4	0	0	0	1	0	0	0	1	0	0	0	94
-0.86	-8.6	1	-8.6	0	0	0	-8.6	0	0	0	1	0	0	0	1	0	0	0	92
-0.85	-8.5	1	-8.5	0	0	0	-8.5	0	0	0	1	0	0	0	1	0	0	0	92
-0.77	-7.7	1	-7.7	0	0	0	-7.7	0	0	0	1	0	0	0	1	0	0	0	92
-0.74	-7.4	1	-7.4	0	0	0	-7.4	0	0	0	1	0	0	0	1	0	0	0	94
-0.65	-6.5	2	-6.5	0	0	0	-6.5	0	0	0	0	1	0	0	1	1	0	0	92
-0.57	-5.7	2	-6.5	0.8	0	0	-5.7	0.8	0	0	0	1	0	0	1	1	0	0	92
-0.23	-2.3	2	-6.5	4.2	0	0	-2.3	4.2	0	0	0	1	0	0	1	1	0	0	92
-0.14	-1.4	2	-6.5	5.1	0	0	-1.4	5.1	0	0	0	1	0	0	1	1	0	0	94
-0.03	-0.3	2	-6.5	6.2	0	0	-0.3	6.2	0	0	0	1	0	0	1	1	0	0	94
0.01	1	3	-6.5	7.5	0	0	1	7.5	0	0	0	0	1	0	1	1	1	0	92
0.21	2.1	3	-6.5	7.5	1.1	0	2.1	8.6	1.1	0	0	0	1	0	1	1	1	0	94
0.23	2.3	3	-6.5	7.5	1.3	0	2.3	8.8	1.3	0	0	0	1	0	1	1	1	0	92
0.52	5.2	3	-6.5	7.5	4.2	0	5.2	11.7	4.2	0	0	0	1	0	1	1	1	0	92
0.71	7.1	3	-6.5	7.5	6.1	0	7.1	13.6	6.1	0	0	0	1	0	1	1	1	0	94
1	10	4	-6.5	7.5	9	0	10	16.5	9	0	0	0	0	1	1	1	1	1	92
1.07	10.7	4	-6.5	7.5	9	0.7	10.7	17.2	9.7	0.7	0	0	0	1	1	1	1	1	92
1.12	11.2	4	-6.5	7.5	9	1.2	11.2	17.7	10.2	1.2	0	0	0	1	1	1	1	1	94
1.55	15.5	4	-6.5	7.5	9	5.5	15.5	22	14.5	5.5	0	0	0	1	1	1	1	1	92
2.5	25	4	-6.5	7.5	9	15	25	31.5	24	15	0	0	0	1	1	1	1	1	93

Table S3.Note 1: The incremental coding scheme used in Experiment 1 is shown in light green. This codes for change in slope in segments 2-4, as well as change in level in segments 2-4. By contrast, an absolute coding scheme is shown with active elements in light blue. While all values in each column contribute to the estimation of coefficients, only the portions that are varying relative to the rest of the column are relevant for interpretation.

Table S3.Note 2: The sum of slope coding across the columns of the absolute scheme sums to the phonotactic probability value (highlighted in pink). Here it is seen that the total slope is decomposed into four distinct slopes, each active in successive segments. This decomposition is not present when coding for change in slope. In the change coding scheme, the 1st segment contains all levels of the predictor and denotes a line active across all levels. The non-zero codes in subsequent segments denote additional lines that deflect (i.e., change) from any prior lines.

Table S3.Note 3: To obtain a predictor range greater than 3.14, average of the two phonotactic probability z-scores used to index phonotactic probability in this experiment (see above section *Non-Word Stimuli for Experiment 1*) was multiplied by 10 (pink highlighted column) to obtain more stable estimation, and more interpretable odds ratios.

Table S3.Note 4: The Freq column shows that except for missing data, each of 20 levels of phonotactic probability is paired with only one nonword. For 47 subjects observed under two conditions (immediate, delayed) there are 94 observations of each nonword.

References

Gould, W. (1993). Linear splines and piecewise linear functions., Stata Technical Bulletin (Vol. 15, pp. 13-17). Available from <http://www.stata.com/products/stb/journals/stb15.pdf>

Lyons, M. (1971). Techniques for using ordinal measures in regression and path analysis. *Sociological Methodology*, 3, 147-171.

Marsh, L. C., & Cormier, D. R. (Eds.). (2002). *Spline Regression Models* (Vol. 137): SAGE Publications.

Panis, C. (1994). Linear splines and piecewise linear functions., Stata Technical Bulletin (Vol. 18, pp. 27-29). Available from <http://www.stata.com/products/stb/journals/stb18.pdf>

UCLA Statistical Consulting Group. How can I run a piecewise regression in Stata? Retrieved 28-Nov, 2012, from <http://www.ats.ucla.edu/stat/stata/faq/piecewise.htm>

Table S4: Segment Coding for Experiment 2

LexDens	Seg	Slope Coding								Level Coding								Freq
		Absolute Segment Slope				Segment Change in Slope				Absolute Segment Level				Segment Change in Level				
		s1	s2	s3	s4	cs1	cs2	cs3	cs4	i1	i2	i3	i4	i1x	i2x	i3x	i4x	
4	1	4	0	0	0	4	0	0	0	1	0	0	0	1	0	0	0	276
5	1	5	0	0	0	5	0	0	0	1	0	0	0	1	0	0	0	414
8	2	8	0	0	0	8	0	0	0	0	1	0	0	1	1	0	0	138
10	2	8	2	0	0	10	2	0	0	0	1	0	0	1	1	0	0	138
11	2	8	3	0	0	11	3	0	0	0	1	0	0	1	1	0	0	414
13	3	8	5	0	0	13	5	0	0	0	0	1	0	1	1	1	0	138
14	3	8	5	1	0	14	6	1	0	0	0	1	0	1	1	1	0	138
15	3	8	5	2	0	15	7	2	0	0	0	1	0	1	1	1	0	138
16	3	8	5	3	0	16	8	3	0	0	0	1	0	1	1	1	0	138
17	3	8	5	4	0	17	9	4	0	0	0	1	0	1	1	1	0	138
18	4	8	5	5	0	18	10	5	0	0	0	0	1	1	1	1	1	276
19	4	8	5	5	1	19	11	6	1	0	0	0	1	1	1	1	1	138
20	4	8	5	5	2	20	12	7	2	0	0	0	1	1	1	1	1	138
24	4	8	5	5	6	24	16	11	6	0	0	0	1	1	1	1	1	138

Table S4.Note 1: The incremental coding scheme employed in Experiment 2 is shown in light green. By comparison, an absolute coding scheme is shown in columns containing light blue.

Table S4.Note 2: Unlike Exp1, four levels of lexical density are paired with more than one nonword. For 69 subjects observed under two conditions (immediate, delayed) there are 138 observations of each nonword. As indicated by the pink highlighting, there were two words with a density of 4, three with a density of 5, three with a density of 11, and two with a density of 18 (See also Table S2).

Figure S1. Nonlinearity in Response to Phonotactic Probability Level in Experiment 1. This figure presents the *predicted* probability of a correct response in the fully-segmented spline mode for Experiment 1 (phonotactic probability). Alternating blue and red markers distinguish each quartile of phonotactic probability. This is the fully segmented model permitting slope and intercept differences across segments. The dashed line is the linear fit line, which provides a visual comparison to the linear model. The visually apparent rise and fall (nonlinearity) of predicted probability of a correct response across low and mid-low phonotactic probability values is contrasted with the dashed line denoting a linear response. By comparison, the relative linearity across segments 3 and 4 is visually apparent. The vertical stratification of predicted values at each level of phonotactic probability denotes the fixed effect of age. The age gap between the 3 and 5 year old children is visually apparent, with better performance by the 5-year-old than 3-year-old children. Lastly, comparison of the top panel (i.e., immediate test) with the bottom panel (i.e., delayed test) shows the fixed effect of time, with better performance at the immediate than delayed test.

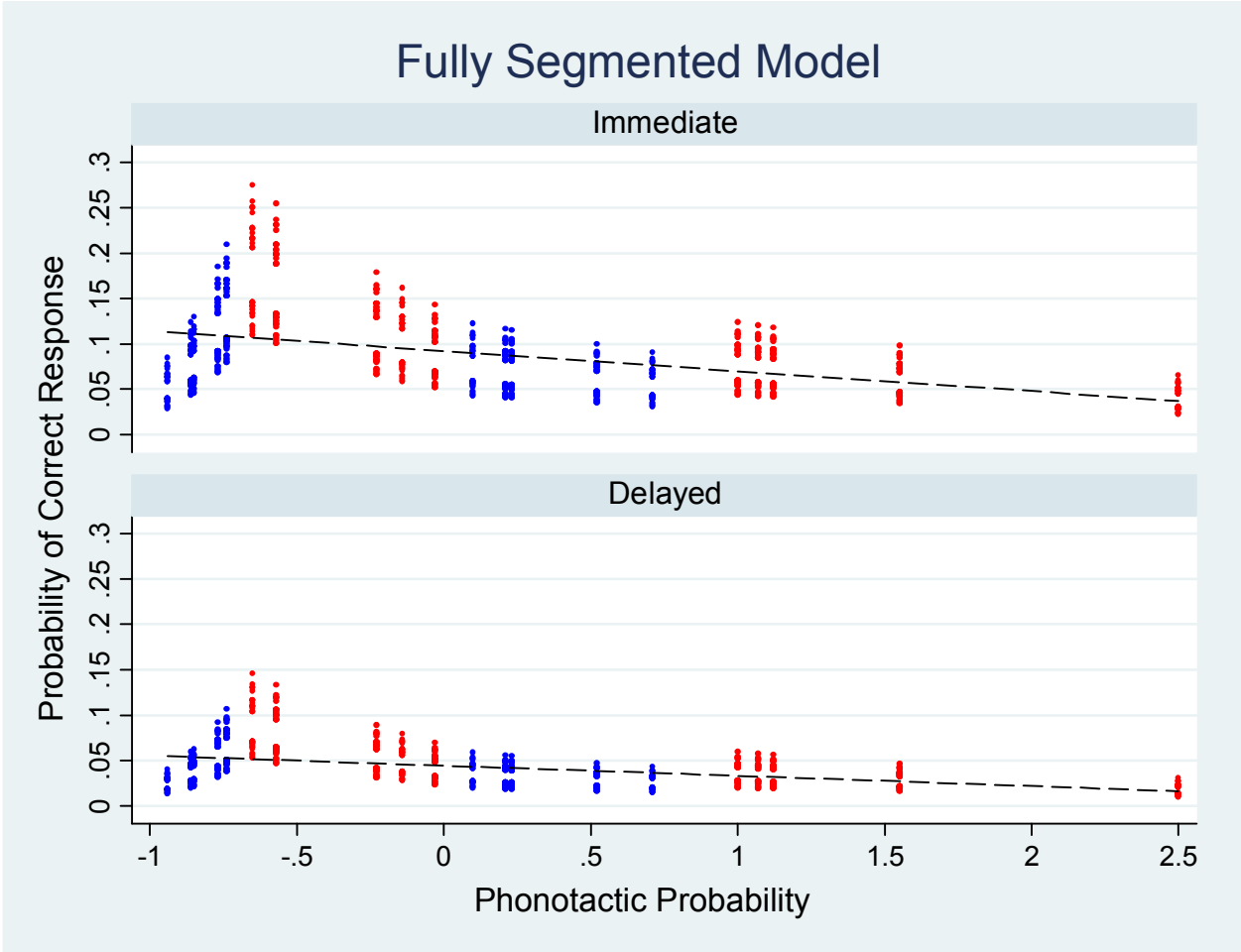


Figure S2. Individual Differences in Probability of Correct Response in Experiment 1. This plot corresponds to a more parsimonious (non-tabulated) phonotactic probability model that coded for significant slope change in the first 3 segments, but did not additionally code for a slope change in the 4th segment, nor any level changes. This is the typical continuous linear spline model. The red lines show the predicted probability of a correct response across fixed effects of phonotactic probability and age; stratification is due to age. The fixed effect for immediate/delayed testing is observed across panels. The blue lines show individual predictions of a correct response across the range of phonotactic probability. The stratification of the blue lines is due to both the fixed effect of the child's age, as well as the random intercept for each child. The age gap between the 3 and 5 year old children is moderately obscured by individual differences of both younger and older subjects.

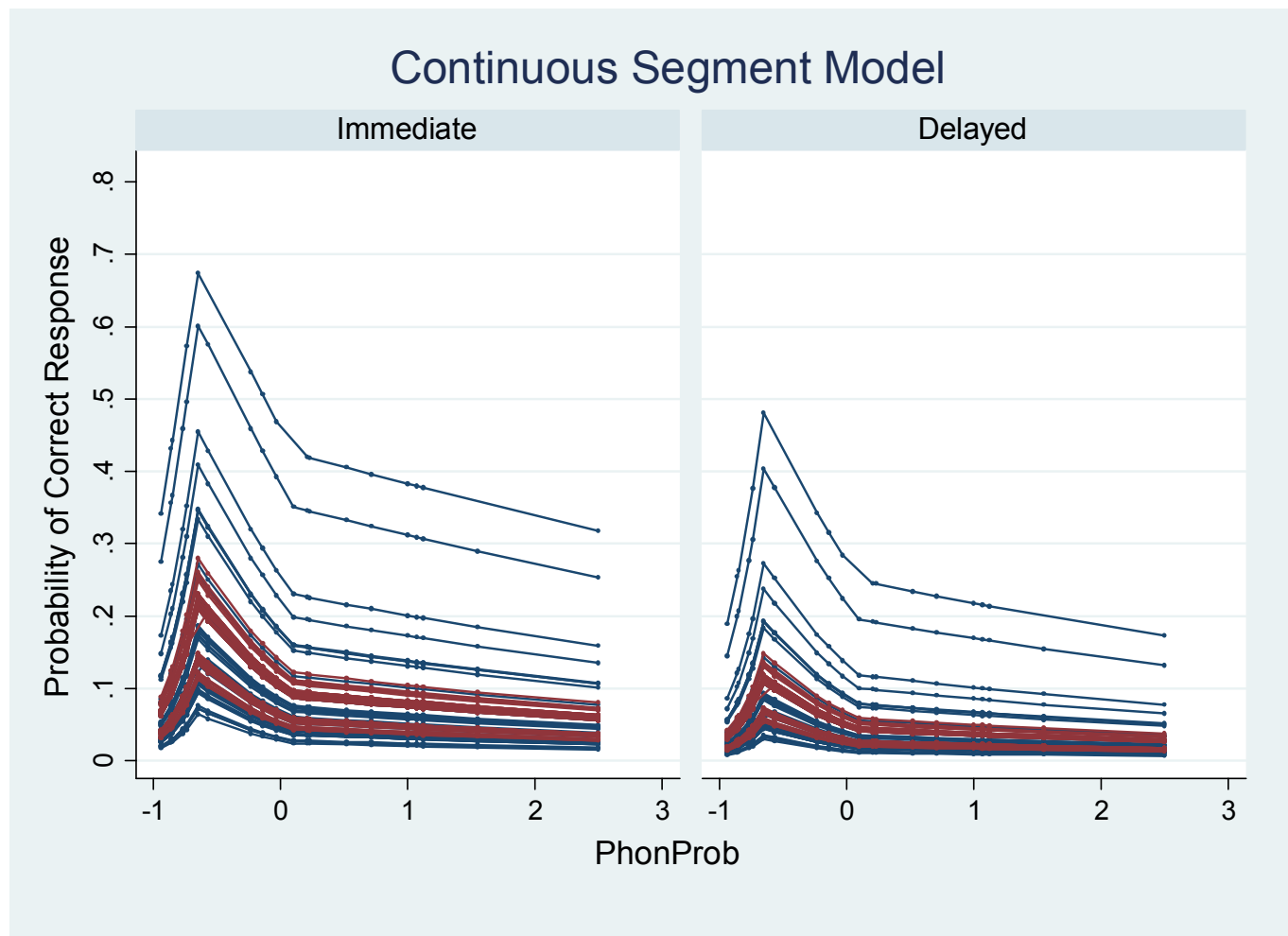


Figure S2.Note 1: Visually the vertical stratification of the fixed effect of age (red) relative to the vertical stratification of the combined effect of age and individual differences (blue) provides further insight into the magnitude of individual differences. The fixed effect of age required a 1.032 increase in the odds of a correct response for each additional month of age. The effect (MOR=2.278) of individual differences (between-subject variance) is considerably larger, as is visually illustrated by the greater range of blue lines.

Figure S3. Individual Differences in Response to Phonotactic Probability Level. This plot contrasts aspects of the random coefficient continuous spline model in Figure S2 with a (non-tabulated) continuous spline model that introduced random coefficients for slopes in segments 1 and 2, thus relaxing the requirement of parallel slopes. Individual colored lines represent the predicted response (fixed + random effects) of an individual child. Same colored lines across panels denote the same child.

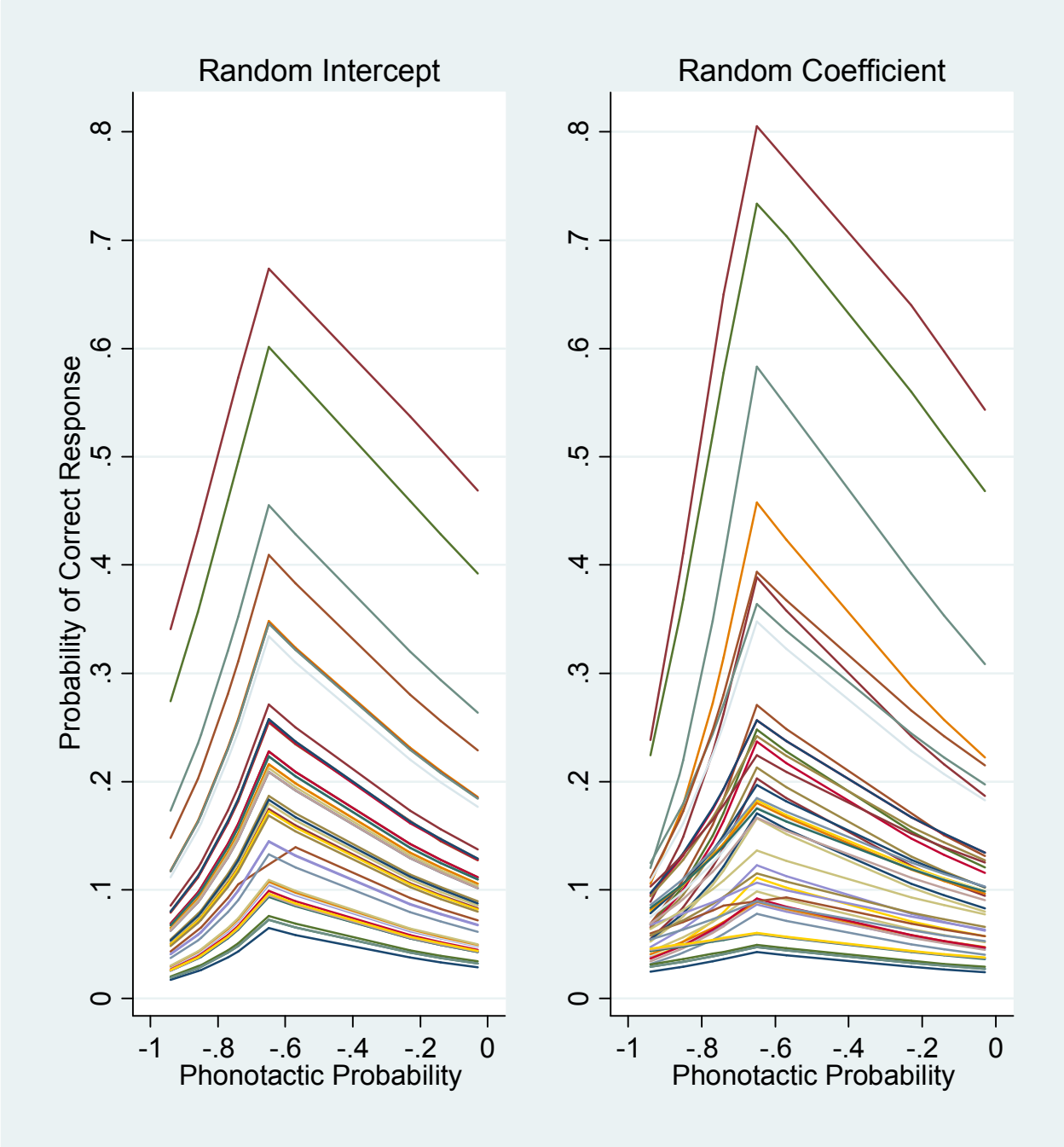
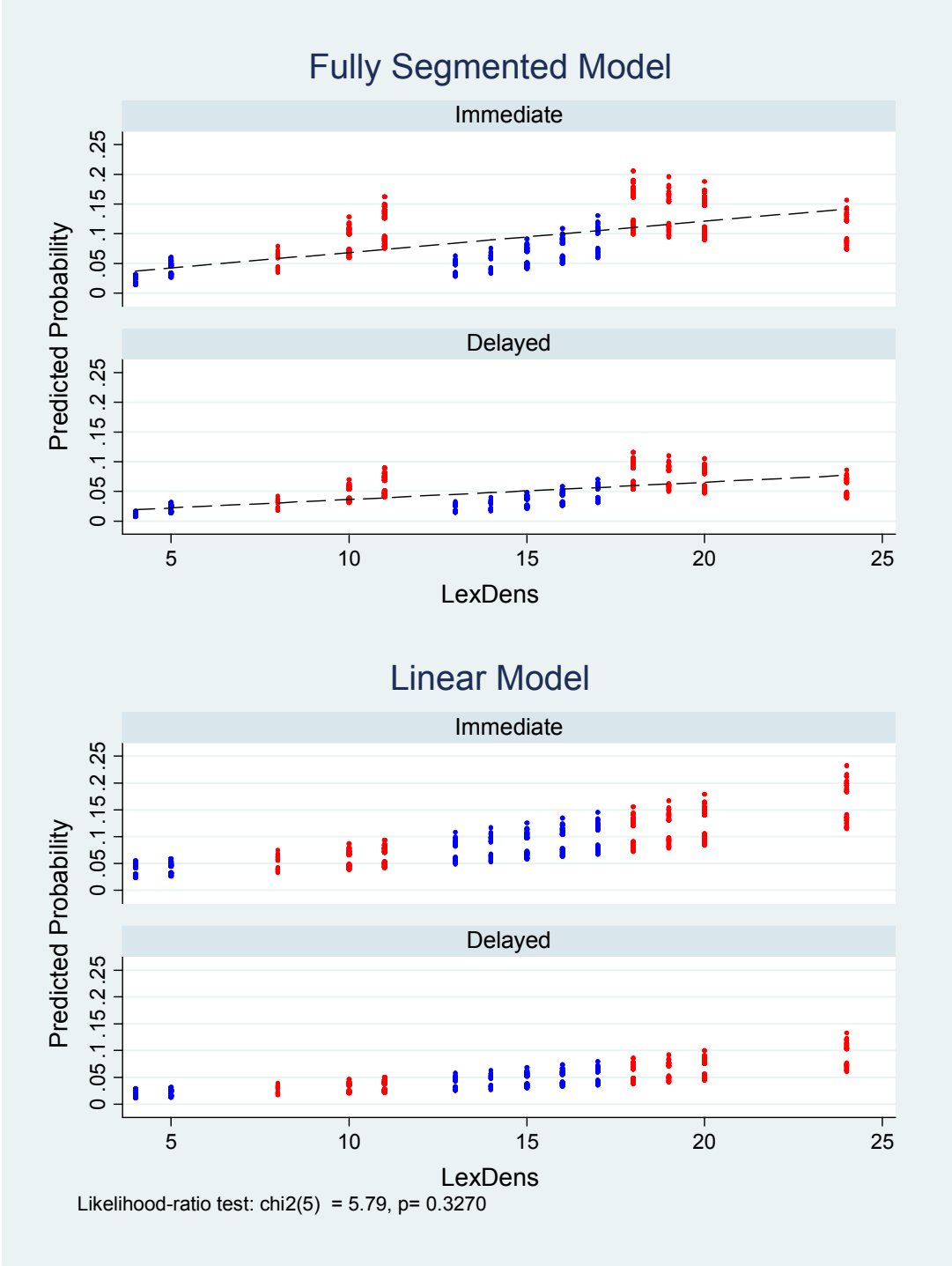


Figure S3.Note 1: The estimated slope variance in segments 1 and 2 was very small with an almost perfect negative correlation, and a strong correlation with the random intercept. Decline corresponded only to prior gain. MOR effects were not computed because it was not clear how to incorporate the

1 covariance of random coefficients into the MOR formula. The respective log likelihood values (RI: -
2 482.4, RC: -479.5) suggested minimal improvement gained by introducing 2 additional variance and
3 3 new covariance parameters. Fixed effects remained virtually unchanged.
4

5 **Figure S3.Note 2:** While relaxing the requirement of parallel slopes did permit higher (and lower)
6 estimates for some children, relatively few lines cross, indicating individual differences are primarily in
7 overall ability levels and not in segment slopes. For example, the three highest and lowest children
8 can be seen to maintain their rank order. One exception is the orange line which rises to a predicted
9 level of just over .2 in the 1st panel, but rises to about .45 in the 2nd panel. However the net effect of
10 such exceptions was negligible on the coefficients for the model's fixed effects of phonotactic
11 probability and age.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure S4. Probability of Correct Response across Lexical Density in Experiment 2. Similar to Figure S1, the upper panel plots predicted probability of a correct response across levels of lexical density, with the dashed line denoting a linear response. This is the fully segmented model permitting slope and intercept differences across segments. Alternating blue and red markers distinguish each quartile of lexical density. The relative linearity across all 4 segments is visually apparent. The vertical stratification of predicted values at each level of lexical density denotes the fixed effect of age. The age gap between the 3 and 5 year old children is visually apparent, with better performance by the 5-year-old than 3-year-old children. Lastly, comparison of the top panel (i.e., immediate test) with the bottom panel (i.e., delayed test) shows the fixed effect of time, with better performance at the immediate than delayed test.



1 The lower panel plots the predictions from the linear model which dropped the fixed slope and level
2 effects for segments 2-4. As indicated by the likelihood ratio test statistic, the fully segmented model
3 does not perform significantly better than the simpler linear model. Thus, the linear model may be
4 preferred because it is more parsimonious. The fixed effect of age and test remain apparent in the
5 figure for this model.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure S5. Individual Differences in Probability of Correct Response in Experiment 2. This plot corresponds to a more parsimonious (non-tabulated) linear lexical density model that used a random intercept for subjects, but dropped the random intercept for items to obtain a more visually approachable plot. This model is analogous to the model in Figure S2 (experiment 1) in that it retains only the fixed effects found to be significant in the fully segmented model. Unlike experiment 1 where several segments did have significant slope changes, in experiment 2 none of the fixed slope or intercept changes were significant. The red lines show the predicted probability of a correct response across fixed effects of neighborhood density and age; stratification is due to age. The fixed effect for immediate/delayed testing is observed across panels. The blue lines show individual predictions of a correct response across the range of neighborhood density. The stratification of the blue lines is due to both the fixed effect of the child's age, as well as the random intercept for each child. The age gap between the 3 and 5 year old children is moderately obscured by individual differences of both younger and older subjects.

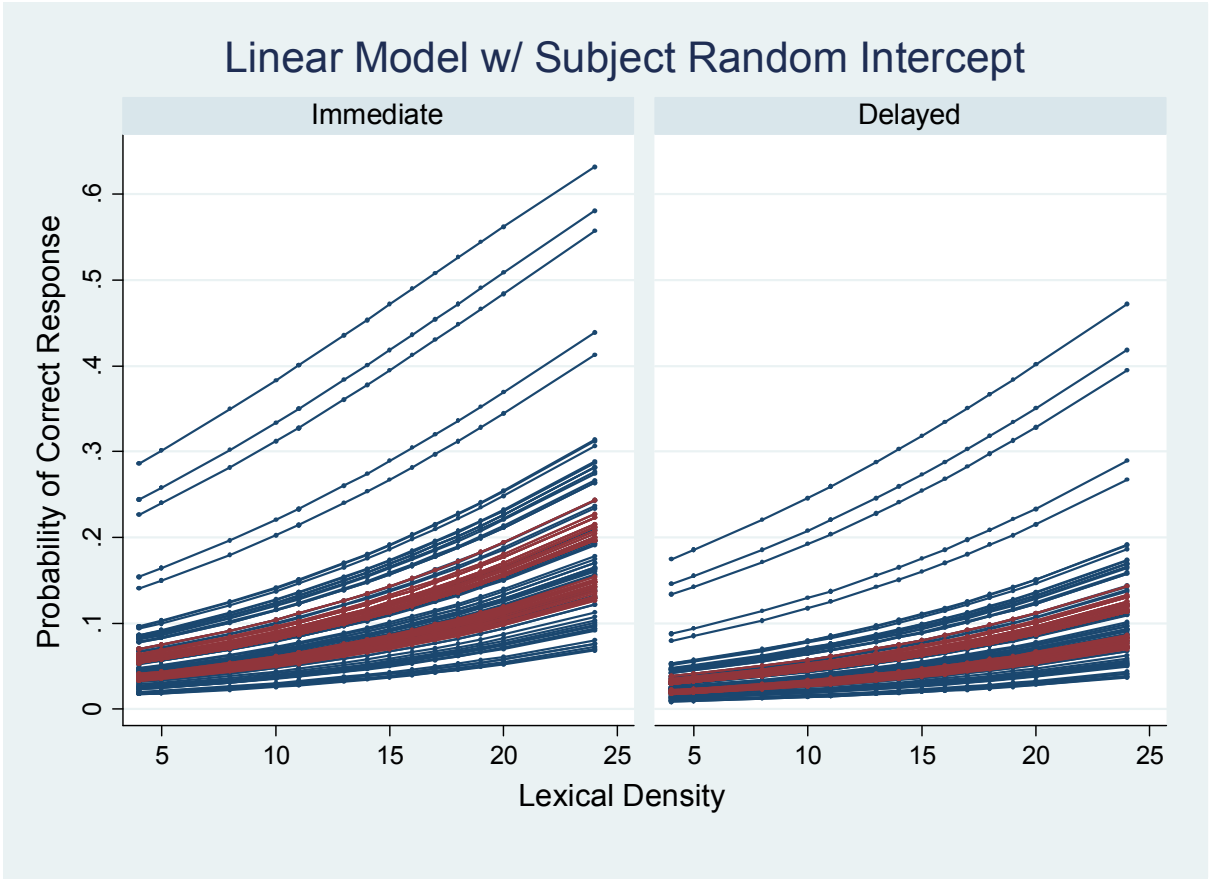


Figure S5.Note 1: Visually the vertical stratification of the fixed effect of age (red) relative to the vertical stratification of the combined effect of age and individual differences (blue) provides further insight into the magnitude of individual differences. The fixed effect of age required a 1.023 increase in the odds of a correct response for each additional month of age. The effect (MOR=2.165) of individual differences (between-subject variance) is considerably larger, as is visually illustrated by the greater range of blue lines.

Figure S5.Note 2: Relaxing the assumption of parallel slopes (i.e., introducing a random coefficient for slope similar to the model in Figure S3) resulted in changes so small as to be visually uninteresting. Individual differences are primarily in level of ability, similar to Experiment 1.